DOCUMENT RESUME

ED 147 313 TM 006 103

AUTHOR Newman, Isadore; Oravecz, Michael T.
TITLE Solutions to the Problem of Disproportionality: A Discussion of the Models.
PUB DATE Apr 77
NOTE 55p.; Paper presented at the Annual Meeting of the American Educational Research Asscciation (61st, New York, New York, April 4-8, 1977); appended computer data may not reproduce well due tc marginal legibility of original document

EDRS PRICE MF-$0.83 HC-$3.50 Plus Postage.
DESCRIPTORS *Analysis of Variance; *Hypothesis Testing; *Mathematical Models; *Research Design; *Research Methodology
IDENTIFIERS *Chi Square Test; *Disproportionality; Type I Error

ABSTRACT

The majcr concern for any research model, whether disproportionate or not, is the research questicn and how well that question is reflected by the model. Three "exact solutions" for disproportional situations, the hierarchial, unadjusted main effects, and fitting constant methods, are discussed in terms of the research questicn that each reflects, and examples demcnstrating the most appropriate situation fcr using each solution are presented. The usefulness of a Chi Square Technique in differentiating between varying degrees of disproportionality and their effects on a Type I error is explained. (Author/HV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED147313

A.E.R.A.
SESSION NO. 28:21

SOLUTIONS TO THE PROBLEM OF DISPROPORTIONALITY: A DISCUSSION OF THE MODELS

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Isadore Newman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ISADORE NEWMAN

MICHAEL T. ORAVECZ

THE UNIVERSITY OF AKRON

PRESENTED AT A.E.R.A.

NEW YORK, NY, APRIL 1977

TM006 103

SOLUTIONS TO THE PROBLEM OF DISPROPORTIONALITY: A DISCUSSION OF THE MODELS*

ISADORE NEWMAN

MICHAEL T. ORAVECZ

The University of Akron

ABSTRACT

There are two major purposes of the paper. The first is to investigate the usefulness of a χ^2 technique in differentiating between varying degrees of disproportionality and their effects on a Type I error. The second purpose is to present and support the position that the major concern for any research model, whether disproportionate or not, is the research question and how well that question is reflected by the model. Three "exact solutions" for disproportional situations, the hierarchial, unadjusted main effects, and fitting constant methods, will also be discussed in terms of the research question that each reflects, and examples will be presented to demonstrate the most appropriate situation for using each solution.

This paper will deal with two main questions. The first will be to identify the effects of different degrees of disproportionality on the nominal level of Type I error and to test the assumption that a χ^2 test can be used to determine when disproportionality is severe enough that corrections are required.

---

*This paper was presented at A.E.R.A., New York, NY, April, 1977.

The second part of this paper will deal with a discussion of some of the more prominent regression approaches ("exact" solutions) used to adjust for disproportionality.

The applied statistician and researcher is plagued with the problem of disproportional cell sizes in factorial experimental designs. This may occur because of mortality in the laboratory animals being used in the experiment; the required number of subjects not available; someone who had agreed to take part in the experiment fails to show up; or the data may represent the proportionality that exists in the "real world." (Disproportionality exists any time the expected values differ from the observed values. Obviously, one can have mild or severe disproportionality. The problem is to determine when the disproportionality is sufficiently severe to require adjustments. Therefore, this paper defines non-significant disproportionality on the bases of a χ^2 test where $\alpha = .25$ and significant disproportionality at $\alpha = .05$.) The effects on factorial designs by significantly disproportional versus non-significantly disproportional cell sizes, as determined by the chi-squared (χ^2) test, have not been studied, or at least reported. Such an investigation was attempted here.

Two studies were made, Study I and Study II. Each study contained four cases with 1,000 experiments per case, with the exception of the fourth case of Study II. In Study I, the main effects in each experiment were not necessarily significantly different. In Study II, the main effects in each experiment were made significantly different.

The following is a discussion of the paradigms used in each study.

The paradigm used in Study I was as follows:

1. Using a computer, 1,000 2 x 2 experiment tables were constructed for each of four cases--equal, proportional, non-significantly disproportional, and significantly disproportional cell sizes. Data points which were uniformly distributed and of equal variance were randomly generated in each of the cells [IBM, 1969].

2. Using computational formulas for sums of squares in the two-factor analysis of variance (ANOVA) with proportional cell frequencies, F-ratios and their corresponding probabilities were computed for each source of variability on each of the four thousand experiments.

3. The result of the analysis on the experiments containing equal frequencies was used as the measure of the actual probability level produced by this research procedure.

4. The probability levels produced by the analysis on the experiments containing significantly disproportional and non-significantly disproportional cell frequencies were compared to each other and to the actual probability level.

The paradigm used in Study II was the same as that for Study I with the exception that a value of five (5) was added to each of the data points of cells one and two so that the row factors were significantly different.

The procedures used in each of the two studies for obtaining the cell sizes, n_i (i = 1, 2, 3, 4), and to determine if the cell sizes were significantly and non-significantly disproportional were as follows:

Case I: Equal n's

A random number between ten and one hundred was generated and this became the size of each cell.

Case II: Proportional n's*

Since the cell sizes are proportional if $\frac{n_1}{n_2} = \frac{n_3}{n_4}$,

n_1, n_2, n_3 were randomly generated and then n_4 was calculated from the equation $n_4 = \frac{n_2 \ n_3}{n_1}$,

where $10 \leq n_i \leq 100$, i = 1, 2, 3. See figure below for the labeling of the cells.

n_1	n_2
n_3	n_4

*Determining n_4 in this manner most often resulted in actual values that n_4 were non-integer numbers which the computer truncated to the integer value. Hence, n_4 differed from the true proportional value an amount x where $0 \leq x < 1$.

Case III: Non-significant disproportional n's

Cell sizes n_1, n_2, and n_3, were randomly generated, and then the proportional cell size n_4 was computed as in Case II. Then, so that the cell sizes were disproportional, a new n_4, n'_4, was calculated from the equation

$$n'_4 = n_4 + x + 1$$

where x is a randomly generated integer between one and ten. To determine if the cell sizes were non-significantly disproportional, a χ^2 test was made ($p \geq \alpha$ .25) where χ^2_c = 1.32--i.e., if $\chi^2_a \leq \chi^2_c$, the n's were judged to be non-significantly disproportional.

Case IV: Significant disproportional n's

Cell sizes n_1, n_2, n_3, and n_4, were randomly generated. To determine if the cells were significantly disproportional, a χ^2 test was made ($p \geq \alpha$ .05) where χ^2_c = 3.841--i.e., if $\chi^2_a > \chi^2_c$, the n's were judged to be significantly disproportional.

The probability of a computed F-ratio was computed by using formulas 26.6.2, 26.6.10, 26.7.8, and 26.2.18 found in Abramowitz and Stegun [1967]. Algebraically, it can be shown that the above formulas combine so that the probability of a F-value for degrees of freedom in the numerator, v_1, and degrees of freedom in the denominator, v_2, is

$$P(F \mid v_1 = 1, v_2)^* = 1 - (1 + c_1x + c_2x^2 + c_3x^3 + c_4x^4)^{-4}$$

---

*P = .95 means that five times out of one hundred an F-value this large or larger would occur due to chance alone.

where $c_1 = .196854$, $c_2 = .115194$, $c_3 = .000344$, $c_4 = .019527$,

$$x = \frac{t\left(1 - \frac{1}{4v_2}\right)}{\sqrt{1 + \frac{t^2}{2v_2}}}, \text{ and } t = \sqrt{F.}$$

TABLE I

PERCENTAGE OF SIGNIFICANT F-RATIOS FOR EACH OF THE FOUR CASES OF STUDY I

SOURCE	EQUAL n's	PROPORTIONAL n's	NON-SIGNIFICANTLY DISPROPORTIONAL n's	SIGNIFICANTLY DISPROPORTIONAL n's
Row Factor	5.60%	4.10%	4.60%	3.60%
Column Factor	4.60%	4.60%	4.40%	3.40%
Interaction	5.40%	5.40%	6.10%	6.90%

TABLE II

PERCENTAGE OF SIGNIFICANT F-RATIOS FOR EACH OF THE FOUR CASES IN STUDY II

SOURCE	EQUAL n's	PROPORTIONAL n's	NON-SIGNIFICANTLY DISPROPORTIONAL n's	SIGNIFICANTLY DISPROPORTIONAL n's
Row Factor	100.00%	100.00%	100.00%	100.00%
Column Factor	6.40%	3.90%	3.50%	**
Interaction	4.80%	6.50%	11.10%	**

** Interaction sums of squares were negative values. The program was written to reject negative F-values. Hence, the process stayed within a loop and no results were available.

The Monte Carlo data supports the position that mild cases of disproportionality versus severe cases of disproportionality, as operationally defined by the χ^2, may be a useful approach in identifying the limits of when one cannot ignore effects of disproportional cells.

As one can see from Table I, when the null hypothesis is true, it appears that disproportionality does not present a problem. However, when the null hypotheses is not true and there is significant disproportionality, as measured by a χ^2 test ($p \leq .05$), the effects of the disproportionality cannot be ignored. In addition, the non-significant disproportionality case, ($.25 < p < 1$) as can be seen from Table II, has more frequently occurring significant interactions than occurred in the equal and proportional cases. Therefore, even in the non-significant case, the interaction may be more severely effected, and one should be sensitive to this possibility.

Obviously, this Monte Carlo study did not investigate the effects of both row and column being significant, row column and interaction being significant, or either row or column and interaction being significant. The reader should be aware that these situations may or may not produce different results than are presented in Tables I and II. However, the investigators feel that at this point, they have produced some supportive evidence for the potential usefulness for using a χ^2 to aid in decision making for determining when it is necessary to use corrections.

THREE REGRESSION SOLUTIONS FOR DISPROPORTIONALITY

When the researchers feel disproportionality is severe enough to be of concern, there are a variety of procedures that he can utilize to attempt to correct for the potential problems. However, before any corrections are applied, one should be sensitive to the underlying assumption that they are making about the population from which their data is drawn, and the investigator must also be very clear about the research question he is interested in asking.

If one had a research project in which the data and variables came from groups that already exist, such as age, intelligence, socio-economic status, etc., and if one was interested in generalizing and predicting back to the group from which the sample came, there is a good possibility that there would be a correlation that would not be spurious between such variables as I.Q. and socio-economic status. In other words, there may be significantly more above average socio-economic status people who have above average I.Q.'s than one would expect by chance. The data were forced to correspond to a balanced design in which there are an equal number of high and low I.Q. people for an equal number of high and low socio-economic status positions, the result of the study and the statistical analysis may allow one to say something that may only be true for that artificially forced relationship and one could not properly generalize to the population in which this

proportionality did not actually exist.

The other side of the coin is if the disproportionality in a research design is an artifact, (it really does not exist in the population) and the disproportionality of vectors is causing a spurious correlation between the variables, then one would have to adjust for this disproportionality and would have to decide which solution of disproportionality would best adjust the data so that it would better reflect the question(s) of interest and the true state of affairs.

To be able to begin to decide upon the correct solution, one has to

a. know something about the theoretical and/or empirical relationship between the variables being studied;

b. know some of the descriptive data about the population one wishes to generalize to in relation to the specific variables being studied;

c. know the specific research question under investigation if one decides an adjustment for disproportionality is needed, then

d. know the underlying assumptions and implications for different adjustment procedures, and

e. know the consequences for using the selected adjustment procedure on the interpretation and generalization of the data.

Many researchers have dealt with items a - c but few have clarified the problems and implications related to

d and e on the above list. The following is a brief discussion of the underlying assumptions of some of the frequently used adjustments for disproportionality. It is hoped that this would aid researchers in being more sensitive to the questions they are really asking when writing models that reflect adjustment by different solutions for the problem of disproportionality.

There has been an exciting and thought provoking debate in the literature as to the most desirable (correct or accurate) procedure for correcting the problems of disproportionality. Most notably, Overall and Spiegel, 1969; Timm and Carlson, 1975; Overall, Spiegel, and Cohen, 1975; Werts and Linn, 1971; Rock, Werts, and Linn, 1976; and Applebaum and Cramer, 1973; have created the interest in the literature.

One basic way of organizing discussions on this topic is to categorize suggestions into two broad groups: approximate solutions or least sum of squares solutions. (This paper will discuss only the least sum of squares solutions.)

Three prominent least sum of squares solutions for disproportionality will be defined.

Solution I is the use of the general linear model to simultaneously adjust for the correlations between the main effects and the main effects with interaction. A symbolic example of this procedure is presented below for a two factorial design.

Model 1 $Y_{kab} = \delta + b_1\alpha_a + b_2\beta_b + b_3\alpha\beta_{ab} + \varepsilon_{kab}$

Model 2 $Y_{kab} = \delta + b_4\beta_b + b_5\alpha\beta_{ab} + \varepsilon_{kab}$

Model 3 $Y_{kab} = \delta + b_6\alpha_a + b_7\alpha\beta_{ab} + \varepsilon_{kab}$

Model 4 $Y_{kab} = \delta + b_8\alpha_a + b_9\beta_b + \varepsilon_{kab}$

Y_{kab} = is the score for subject k in row a and column b

δ = is the grand $\bar{X}$

α_a = is the effect for row "a"

β_b = is the effect for column "b"

$\alpha\beta_{ab}$ = is the interaction effect for the row "a" and column "b"

ε_{kab} = is the error term for each subject

$b, \ldots b_n$ are partial regression coefficients

Adjustment for Solution #1

Adjustment for A main effects test Model 1 against Model 2

Adjustment for B main effects test Model 1 against Model 3

Adjustment for A*B effects test Model 1 against Model 4

(see also Figure I, P. 19)

Solution 2 adjusts each main effect in terms of the other main effects. The interaction was adjusted for by all main effects. (This is the same as in Solution 1). The following is a symbolic representation of this solution: (see also Figure II, P. 20)

Adjustment for Solution #2

Model 4 $Y_{kab} = \delta + b_{10}\alpha_a + b_{11}\beta_b + \varepsilon_{kab}$

Model 5 $Y_{kab} = \delta + b_{12}\beta_b + \varepsilon_{kab}$

Model 6 $Y_{kab} = \delta + b_{13}\alpha_a + \varepsilon_{kab}$

Adjustment for A main effects test Model 4 against Model 5

Adjustment for B main effects test Model 4 against Model 6

Adjustment for AB interaction effects test Model 4 against Model 1

Solution 3 assumes an apriori ordering of the importance of the variables under investigation. The apriori ordering decides which variables one will allow to account for as much variance as possible by themselves. The following is a symbolic representation of Solution 3, assuming the researcher considers the A main effects most important, B main effects second, and the interaction least important.

(Many researchers feel that it is unlikely that most investigators will be able to order the importance of their variable. However, we believe this judgment can be made by a competent researcher who is aware of the underlying constructs and theories he is dealing with) (see also Figure III, P. 21)

Admustment for Solution #3

Model 7 $Y_{kab} = \delta + b_{14}\alpha_a + \varepsilon_{kab}$

Model 8 $Y_{kab} = \delta + \varepsilon_{kab}$

Model 9 $Y_{kab} = \delta + b_{15}\alpha_a + b_{16}\beta_b + \varepsilon_{kab}$

Adjustment for A main effects test Model 7 against Model 8

Adjustment for B main effects test Model 9 against Model 7

Adjustment for AB interaction test Model 1 against Model 9

Marks (1974) presents a mathematical proof that one has to have to A main effects and B main effects in the full and restricted models to test for AB interaction. It seems that this is true when dealing with traditional analysis of variance and catagorical variables. That is, the A main effects and B main effects must be fitted first before interaction can be tested. However, this does not seem to be necessarily true when dealing with continuous variables. To the best of our knowledge, this has never been investigated.

Most researchers, specifically the highly mathematically oriented statisticians, tend to evaluate the accuracy of the above three least square solutions for disproportionality in terms of some arbitrary mathematical matrices often based upon unfounded underlying assumptions. A more logical method for deciding upon which solution is most appropriate in a specific case would be to be aware of the underlying assumptions for each of the structural models of each solution and assumptions about the relationship between the variables being studied.

In a 2 x 2 orthogonal design, the correlation between the A main effect, B main effect, and the AB interaction is zero. Therefore, the effects of the A main effect in predicting the criterion can be totally attributed to the A main effect; similarly for the B main effect and the AB interaction. However, when the design is non orthogonal (disproportional n's) correlations between the A main effects, B main effects, and AB interaction are likely. Therefore, the effect of A in predicting the criterion is not necessarily independent of the B main effect and AB interaction.

Each of the three least square solutions make different assumptions about the meaningfulness and "usefullness" of the correlations between the A main effect, B main effect, and AB interaction.

Solution 1, for example, when testing the A main effect, assumes the correlation between A and B and the AB interaction is of an accidental nature, and therefore should not be considered (Rock, et al., 1976). This solution is most likely to be prefered when one can assume that the missing subjects producing disproportionality were random. If one is unable to make this assumption, then it would be inappropriate to use Solution 1, (which may be the case most frequently).

Solution 2 assumes that there is no correlation between the A and B main effects in the population. Therefore, the correlation between A and B in the sample is a function of disproportionality and not representative of the population. Solution 2 then attempts to adjust for this correlation.

However, Solution 2 assumes that the correlations between the main effects and the interaction, which results from the disproportionality, are not spurious and are characteristic of the population. Therefore, it does not attempt to adjust for this correlation.

If one cannot assume that the correlations between the A and B main effects, due to disproportionality, are due to chance, than Solution 2 would be an inappropriate correction.

Solution 3 requires an apriori ordering of the importance of each variable. Let us assume that the apriori ordering are: A main effects, B main effects, AB interaction, respectively.

When testing for the A main effects, Solution 3 assumes that the correlation between the A main effects, B main effects, and A main effects with the AB interaction is representative of the relationship between these variables in the population and therefore relevant to the research question of interest. That is to say, the relationship between these variables are not artifacts of disproportionality.

However, when testing for B main effects, Solution 3 assumes that the correlation between the A main effect and B main effect is not relevant and is therefore an artifact of unequal N's. It also assumes that the correlation between the A main effects and the interaction, and the B main effects and the interaction is not due to just a chance relationship caused by disproportionality.

The key to the most appropriate selection of these three solutions is not as much a statistical concern (mathematical) but rather a comprehensive understanding of the data so that one can more accurately speculate on which correlations between the variables are likely to be meaningful or not meaningful.

FIGURE 1

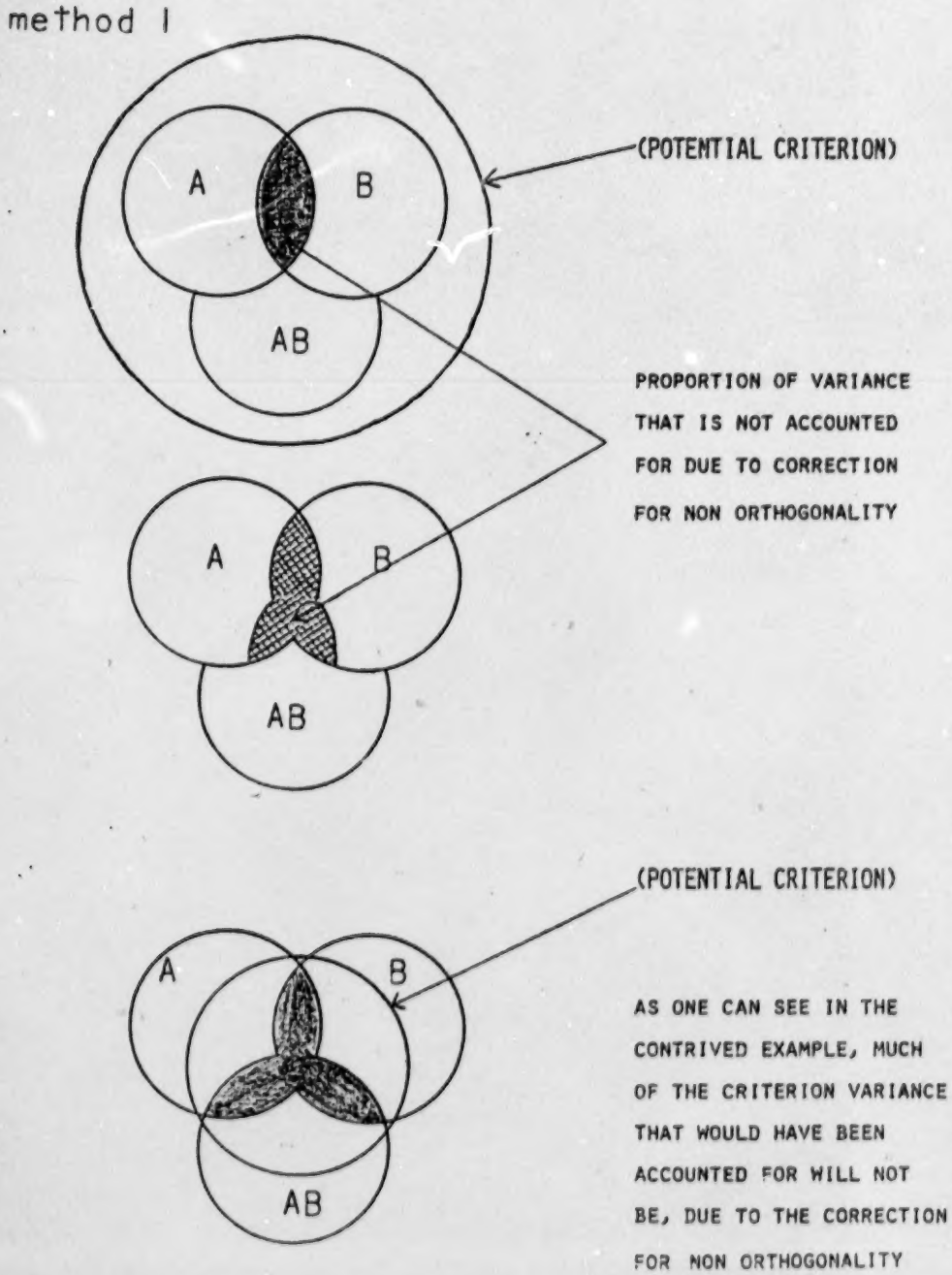

method 1
A
B
AB
(POTENTIAL CRITERION)
PROPORTION OF VARIANCE
THAT IS NOT ACCOUNTED
FOR DUE TO CORRECTION
FOR NON ORTHOGONALITY
A
B
AB
(POTENTIAL CRITERION)
A
B
AB
AS ONE CAN SEE IN THE
CONTRIVED EXAMPLE, MUCH
OF THE CRITERION VARIANCE
THAT WOULD HAVE BEEN
ACCOUNTED FOR WILL NOT
BE, DUE TO THE CORRECTION
FOR NON ORTHOGONALITY

FIGURE 11

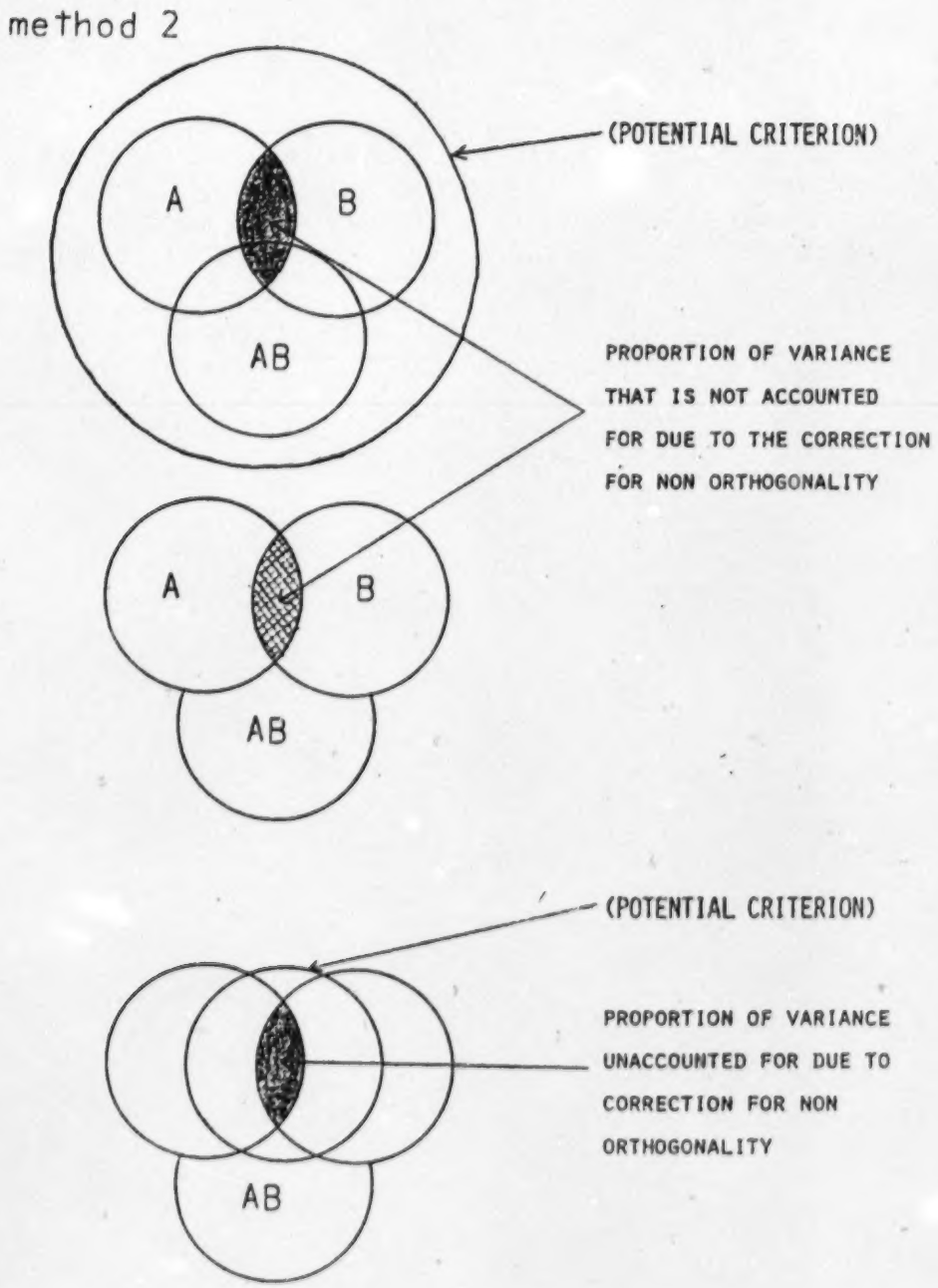
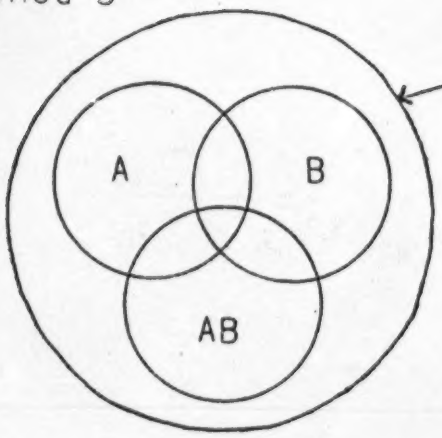

method 2
(POTENTIAL CRITERION)
A
B
AB
PROPORTION OF VARIANCE
THAT IS NOT ACCOUNTED
FOR DUE TO THE CORRECTION
FOR NON ORTHOGONALITY
A
B
AB
(POTENTIAL CRITERION)
PROPORTION OF VARIANCE
UNACCOUNTED FOR DUE TO
CORRECTION FOR NON
ORTHOGONALITY
AB

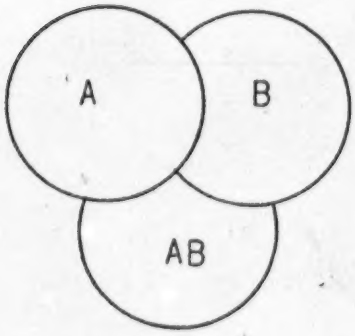
FIGURE 1i1

method 3



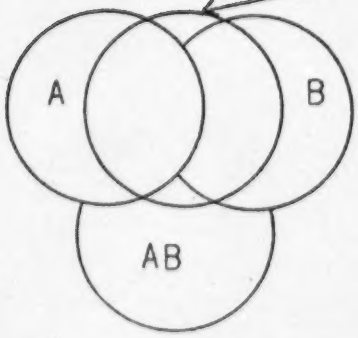
(POTENTIAL CRITERION)

ASSUMING THAT THE ORDERING OF THE VARIABLES OF INTEREST ARE A, B, AB, RESPECTIVELY.



THERE IS NO VARIANCE UNACCOUNTED FOR DUE TO THIS CORRECTION FOR NON ORTHOGONALITY

(POTENTIAL CRITERION)



SOME ADDITIONAL CONSIDERATIONS

Since the three solutions previously discussed are only defined for use with fixed designs, it is initially important to define the basic statistical models of fixed randomized and mixed effects. The fixed effect model is a model in which all the variables have levels which are fixed. That is, the levels of a variable that is to be investigated are determined prior to the investigation. These categorical (high, medium low, etc.,) determinations set the limits to which the investigator can generalize his results.

Randomized effect is a design in which the researcher randomly selects the level of variable that is to be investigated from an infinitely large number of possible levels. This allows the researcher to generalize his results to the entire range of the variable being investigated (that is to the extent that the levels were truly randomly selected and are representative).

The mixed effect model has at least one fixed effect variable and one random effect variable.

Most research conducted is on fixed effect models. This model has a great deal research behind it which indicated that the model is very robust. This means, it is little effected by violations of its assumptions of normality, homogeneity,etc. However, this robustness is not true of the randomized effects model. This model is very sensitive to violations of normality and homogeneity of

variance. Most interestingly, in terms of the purpose of this paper, is that the assumptions of equal N's is absolutely essential for deriving correct error terms for randomized effect designs.[1]

Assuming that one is dealing only with a fixed effect design, one alternative is to consider the correlation between the variables, due to the disproportionality as accurately existing in the population one wishes to generalize to. Then, the researcher may choose not to correct for the disproportionality. The problems that arise with this, is that one cannot attribute the variance accounted for to a particular variable. (This is the dilema of much ex post facto research.)

In conclusion, as can be seen from the discussion in Part II of this paper, it appears obvious that it would be inappropriate to attempt to decide on a correction for disproportionality by using Monte Carlo Studies. Since the appropriateness of a solution depends upon the correlations as one can assume to exist between the variables in the population,and this can change from population to population or with varing theoretical positions, it is much more relevant to try to understand these relationships and base your solution on this understanding than to base

---

[1]To our knowledge, there has been no research on methods for correcting unequal N's in randomized designs.

your decision on the outcome of generated data in which the relationships may be totally unrelated to the relationships in the population of interest.

A highly detailed discussion of different solutions to the hypothesis being tested can be found in Searle (1971). There exists problems that go beyond the scope of this paper such as what occurs when interactions are included in the model and one or more cells equal zero; Searle (1971), Marks (197), and Williams (1977), deal with the problems of full versus non-full ranked matrices; Williams (1977) indicates potentially different effects on solutions due to the effects of different coding procedures. All of these questions are of interest. owever, once again the key is to understand your research questions and the assumptions about the relationships that exist among your variables as reflected by your models.

REFERENCES

Abramowitz, M., & Stegun, I. A. (Eds). Handbook of mathematical functions with formulas, graphs, and mathematical tables. (National Bureau of Standars Applied Mathematics Series 55) Washington, D.C.: United States Government Printing Office, 1967.

Applebaum, M. I. and Cramer, E. M. Some problems in the nonorthogonal analysis of variance. Psychological Bulletin, 1974, 81, 335-343.

Carlson, J. E., & Timm, N. H. Analysis of nonorthogonal fixed-effeccts designs. Psychological Bulletin, 1974, 81 (9), 563-570.

Cohen, J., & Cohen, P. "Multiple regression as a general data analytic system," Psychological Bulletin, 1968 70, pp. 426-443.

Dalton, Starrett. Regression approaches and approximate solutions to analysis of variance with disproportionality varied. Multiple Linear Regression Viewpoints. Vol. 7, 2, Jan., 1977.

Darlington, R. B. Multiple regression in research and practice. Psychological Bulletin, 1968, 69, 166-168.

Glass, G. V., & Stanley, J. C. Statistical methods in education and psychology. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

I.B.M. Random number generation and testing. White Plains, N.Y.: I.B.M., Technical Publications Dept., 1969.

Marks, Edmond. Analysis of variance techniques for unbalanced data. Review of Educational Research. Vol. 44, 3, 1974.

Overall, J. E., & Spiegel, D. K. Concerning least squares analysis of experimental data. Psychological Bulletin, 1969, 72, 311-322.

Overall, J. E., Spiegel, D. K., & Cohen, J. Equivalence of orthogonal and nonorthogonal analysis of variance. Psychological Bulletin, 1975, 82, 182-186.

Rock, Werts, & Linn. Structural equations as an aid in the interpretation of the non-orthogonal analysis of variance. Multivariate Behavioral Research, 1976, 443-448.

Speed, F. M. & Hocking R. R. The use of the R( )-notation with unbalanced data. The American Statistican. Feb., 1976, Vol. 30, 1.

Timm, N. H., & Carlson, J. E. Analysis of variance through full rank models. Monograph issued by Multivariate Behavioral Research, 1975.

Williams, J. D. Two way fixed effects analysis of variance with disproportionate cell frequencies. Multivariate Behavioral Research, 1972, 67-83.

Williams, J. D. "Full rank and non full rank models with contrast and binary coding systems for a two-way disproportionate cell frequency analysis." (In press)

APPENDIX

28

```
C ON110 ORAVECZ-NEWMAN HYPOTHESIS - EQUAL N'S CASE
C MODIFIED FROM ON100  3-2-76  WD WHEATON
C DEFINITION - 2X2 CONTINGENCY TABLE
C   CELL(1,1) EQU CELL(1)
C   CELL(1,2) EQU CELL(2)
C   CELL(2,1) EQU CELL(3)
C   CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),PPCT(6)
      INTEGER*4 N(4),NSIG(3),DATE(4)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
200   FORMAT('1',35X,'ON110  ORAVECZ & NEWMAN EQUAL N''S CASE  '
     1,4A2,T123,'PAGE',I3//
     2'          CELL SIZE RATIOS                    CELL POPULATION SIZE
     3          F VALUES                             PROBABILITIES'/
     4' ------------------------------------   ------------------------
     5 ------------------------------------   ------------------------'/
     6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
     7'    ROW        COL        TABLE        ROW      COL      TABLE'/
     8' -------  -------  -------  -------  -----  -----  -----  -----
     9' ---------  ---------  ---------  -------  -------  -------')
201   FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5)
202   FORMAT('1ON110  ORAVECZ & NEWMAN EQUAL N''S CASE  ',4A2///
     111X,'PERCENT OF CASES'/
     210X,18('-')/
     312X,'SIG      NON-SIG'/
     410X,'-------   -------'/
     5' ROW  ',2F11.2//
     6' COL  ',2F11.2//
     7' TABLE',2F11.2///' TOTAL CASES RUN =',I6)
C GET CURRENT DATE
      CALL TDATE(DATE)
C CONSTANT NEEDED LATER  D1=2./9.=0.222222
      D1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=0.196854
      CC2=0.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER OF CASES TO RUN
      NC=1000
C NUMBER OF NON-SIGNIFICANT CASES - ROW,COL & TABLE
      NSIG(1)=0
      NSIG(2)=0
      NSIG(3)=0
C PAGE & LINE CNTS
      LINE=100
      IPAGE=1
      IY=9
C LOOP OVER ALL CASES
      DO 306 K=1,NC
C GENERATE CELL SIZE  (.GT. 10)
300   IX=IY
      CALL RANDU(IX,IY,PP)
      NN=PP*100.
      IF(NN.LT.10) GO TO 300
C SET ALL CELLS EQUAL & SET CELL PROBABILITIES TO ONE
      DO 301 I=1,4
```

```
      N(I)=N
      P(I)=0.25
  301 CONTINUE
      NSUM=4*N
C GENERATE MEMBERS OF EACH CELL & CELL SUM
  310 TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
       SUMS(I)=0.0
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
  304 CONTINUE
      TSUMS=TSUMS+SUMS(I)
  303 CONTINUE
C CALC ROW,COL,& TABLE SUMS
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
      FC=SSC/SERRM
      FRC=SSRC/SERRM
C IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 310
      IF(FC.LT.0.0) GO TO 310
      IF(FRC.LT.0.0) GO TO 310
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      D2=D1/DF
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIGNIFICANT CASES - ROW,COL & TABLE
      IF(PF(I).LE.0.95) NSIG(I)=NSIG(I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LT.55) GO TO 307
      WRITE(6,220) DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(6,231)P,N,F,PF
      LINE=LINE+1
```

```
306  CONTINUE
C CALC SIGNIFICANT CASES,PERCENTAGES, & PRINT SUMMARY
      ANC=NC
      DO 308 L=1,5,2
      N=L/2+1
      LSIG=NSIG(N)
      MSIG=NC-LSIG
      PRCT(L)=FLOAT(MSIG)*100/ANC
      PRCT(L+1)=FLOAT(LSIG)*100/ANC
308  CONTINUE
      WRITE(6,202)DATE,PRCT,NC
      CALL EXIT
      END
```

31

```
C ON110 ORAVECZ-NEWMAN HYPOTHESIS - EQUAL N'S CASE
C MODIFIED FROM ON100  3-2-76  WD WHEATON  ROW FACTOR SIGNIFICANT
C DEFINITION - 2X2 CONTINGENCY TABLE
C   CELL(1,1) EQU CELL(1)
C   CELL(1,2) EQU CELL(2)
C   CELL(2,1) EQU CELL(3)
C   CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),PRCT(6)
      INTEGER*4 N(4),NSIG(3),DATE(2)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
 200  FORMAT('1',35X,'ON110  ORAVECZ & NEWMAN EQUAL N''S CASE  '
     1,4A2,T123,'PAGE',I3//
     2'          CELL SIZE RATIOS                    CELL POPULATION SIZE
     3               F VALUES                          PROBABILITIES'/
     4' ------------------------------------  ------------------------
     5 ------------------------------------  ------------------------'/
     6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
     7    ROW          COL          TABLE      ROW       COL       TABLE'/
     8' -------  -------  -------  -------  -----  -----  -----  -----
     9 ---------  ---------  ---------  -------  -------  -------')
 201  FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5)
 202  FORMAT('1ON110  ORAVECZ & NEWMAN EQUAL N''S CASE  ',2A4///
     111X,'PERCENT OF CASES'/
     210X,18('-')//
     312X,'SIG       NON-SIG'/
     410X,'-------    -------'/
     5' ROW  ',2F11.2//
     6' COL  ',2F11.2//
     7' TABLE',2F11.2///' TOTAL CASES RUN =',I6)
C GET CURRENT DATE
      CALL TDATE(1,DATE)
C CONSTANT NEEDED LATER  D1=2./9.=0.222222
      D1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=0.196854
      CC2=0.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER OF CASES TO RUN
      NC=1000
C NUMBER OF NON-SIGNIFICANT CASES - ROW,COL & TABLE
      NSIG(1)=0
      NSIG(2)=0
      NSIG(3)=0
C PAGE & LINE CNTS
      LINE=100
      IPAGE=1
      IY=9
C LOOP OVER ALL CASES
      DO 306 K=1,NC
C GENERATE CELL SIZE  (.GT. 10)
 300  IX=IY
      CALL RANDU(IX,IY,PP)
      NN=PP*100.
      IF(NN.LT.10) GO TO 300
C SET ALL CELLS EQUAL & SET CELL PROBABILITIES TO ONE
      DO 301 I=1,4
```

32

```
      N(I)=NN
      P(I)=0.25
  301 CONTINUE
      NSUM=4*NN
C GENERATE MEMBERS OF EACH CELL & CELL SUM
  310 TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
       SUMS(I)=0.0
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
      IF(I.LE.2) A=A+5
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
  304 CONTINUE
      TSUMS=TSUMS+SUMS(I)
  303 CONTINUE
C CALC ROW,COL,& TABLE SUMS
C    CALC ROW,COL,& TABLE SUMS SO ROW FACTOR SIGNIF
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
      FC=SSC/SERRM
      FRC=SSRC/SERRM
C IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 310
      IF(FC.LT.0.0) GO TO 310
      IF(FRC.LT.0.0) GO TO 310
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      C2=D1/DF
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIGNIFICANT CASES - ROW,COL & TABLE
      IF(PF(I).LE.0.95) NSIG(I)=NSIG(I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(6,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
```

```
 307  WRITE(6,201)P,N,F,PF
      LINE=LINE+1
 306  CONTINUE
C CALC SIGNIFICANT CASES,PERCENTAGES, & PRINT SUMMARY
      ANC=NC
      DO 308 L=1,6,2
      M=L/2+1
      LSIG=NSIG(M)
      MSIG=NC-LSIG
      PRCT(L)=FLOAT(MSIG)*100/ANC
      PRCT(L+1)=FLOAT(LSIG)*100/ANC
 308  CONTINUE
      WRITE(6,202)DATE,PRCT,NC
      STOP
      END
```

```
ON100 CRAVECZ-NEWMAN HYPOTHESIS - PROPORTIONAL CASE
WRITTEN 2-11-76  WD WHEATON          ROW FACTOR NOT
DEFINITION - 2X2 CONTINGENCY TABLE   NECESSARILY SIGNIFICANT
 CELL(1,1) EQU CELL(1)
 CELL(1,2) EQU CELL(2)
 CELL(2,1) EQU CELL(3)
 CELL(2,2) EQU CELL(4)
     REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),PRCT(6)
     INTEGER*4 N(4),NSIG(3),DATE(4)
     EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
    1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
    2(F(1),FR),(F(2),FC),(F(3),FRC)
200  FORMAT('1',35X,'ON100  CRAVECZ & NEWMAN PROPORTIONAL CELL SIZES  '
    1,4A2,T123,'PAGE',I3//
    2' ------------ CELL SIZE RATIOS ------------  ------------ CELL POPULATION SIZE
    3 ------------ F VALUES ------------  ------------ PROBABILITIES'/
    4' ------------------------------------  ------------------------------
    5 ------------------------------------  ------------------------------'/
    6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
    7     ROW        COL        TABLE        ROW      COL      TABLE'/
    8'-------- --------- --------- ---------  ----- ----- ----- -----
    9 -------- --------- --------- ---------  ------- ------- -------')
201  FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5)
202  FORMAT('1ON100  CRAVECZ & NEWMAN PROPORTIONAL CELL SIZES  ',4A2///
    111X,'PERCENT OF CASES'/
    210X,18('-')/
    312X,'SIG      NON-SIG'/
    410X,'-------   -------'/
    5' ROW  ',2F11.2//
    6' COL  ',2F11.2//
    7' TABLE',2F11.2///' TOTAL CASES RUN =',I6)
 GET CURRENT DATE
     CALL TDATE(DATE)
 CONSTANT NEEDED LATER  D1=2./9.=0.222222
     D1=0.222222
 PROBABILITY EXPANSION CONSTANTS
     CC1=0.196854
     CC2=0.115194
     CC3=0.000344
     CC4=0.019527
 NUMBER OF CASES TO RUN
     NC=1000
 NUMBER OF NON-SIGNIFICANT CASES - ROW,COL & TABLE
     NSIG(1)=0
     NSIG(2)=0
     NSIG(3)=0
 PAGE & LINE CNTS
     LINE=100
     IPAGE=1
     IY=9
 LOOP OVER ALL CASES
     DO 306 K=1,NC
 GENERATE CELL SIZE  (.GT. 10)
300  IX=IY
     CALL RANDU(IX,IY,PP)
     NN=PP*100.
     IF(NN.LT.10) GO TO 300
 CALC CELL SIZES
     NSUM=0
```

```
      DO 301 I=1,3
02    IX=IY
      CALL RANDU(IX,IY,A)
      NN=A*100.
      IF(NN.LT.10) GO TO 302
      N(I)=NN
      NSUM=NSUM+NN
01    CONTINUE
      N4=N2*N3/N1
IF N4 IS LESS THAN 6 DROP CASE
      IF(N4.LT.6) GO TO 300
      NSUM=NSUM+N4
      SUMN=NSUM
      DO 309 I=1,4
      AN=N(I)
      P(I)=AN/SUMN
309   CONTINUE
GENERATE MEMBERS OF EACH CELL & CELL SUM
310   TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
      SUMS(I)=0.0
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
304   CONTINUE
      TSUMS=TSUMS+SUMS(I)
303   CONTINUE
CALC ROW,COL,& TABLE SUMS
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
      FC=SSC/SERRM
      FRC=SSRC/SERRM
IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 310
      IF(FC.LT.0.0) GO TO 310
      IF(FRC.LT.0.0) GO TO 310
CALC PROBABILITIES FOR ROWS,COLS & TABLE
      B2=B1/DF
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
```

```
      X2=1.+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIGNIFICANT CASES - ROW,COL & TABLE
      IF(PF(I).LE.0.95) NSIG(I)=NSIG(I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(6,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(6,201)P,N,F,PF
      LINE=LINE+1
  306 CONTINUE
C CALC SIGNIFICANT CASES,PERCENTAGES, & PRINT SUMMARY
      ANC=NC
      DO 308 L=1,6,2
      M=L/2+1
      LSIG=NSIG(M)
      MSIG=NC-LSIG
      PRCT(L)=FLOAT(MSIG)*100/ANC
      PRCT(L+1)=FLOAT(LSIG)*100/ANC
  308 CONTINUE
      WRITE(6,202)DATE,PRCT,NC
      CALL EXIT
      END
```

```
LEVEL  21                    MAIN                    DATE = 76166          11/43/20

C ON100 ORAVECZ-NEWMAN HYPOTHESIS - PROPORTIONAL CASE
C WRITTEN 2-11-76  WD WHEATON                ROWFACTOR SIGNIFICANT
C DEFINITION - 2X2 CONTINGENCY TABLE
C  CELL(1,1) EQU CELL(1)
C  CELL(1,2) EQU CELL(2)
C  CELL(2,1) EQU CELL(3)
C  CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),PRCT(6)
      INTEGER*4 N(4),NSIG(3),DATE(2)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
 200  FORMAT('1',35X,'ON100  ORAVECZ & NEWMAN PROPORTIONAL CELL SIZES  '
     1,4A2,T123,'PAGE',I3//
     2'           CELL SIZE RATIOS                    CELL POPULATION SIZE
     3                F VALUES                          PROBABILITIES'/
     4' ---------------------------------  ----------------------------
     5 ---------------------------------  ----------------------------'/
     6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
     7     ROW          COL          TABLE       ROW      COL      TABLE'/
     8' --------  --------  --------  --------  -----  -----  -----  -----
     9 ---------  ---------  ---------  --------  -------  -------  -------')
 201  FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5)
 202  FORMAT('1ON100  ORAVECZ & NEWMAN PROPORTIONAL CELL SIZES  ',2A4///
     111X,'PERCENT OF CASES'/
     210X,18('-')/
     312X,'SIG      NON-SIG'/
     410X,'-------    -------'/
     5' ROW  ',2F11.2//
     6' COL  ',2F11.2//
     7' TABLE',2F11.2///' TOTAL CASES RUN =',I6)
C GET CURRENT DATE
      CALL TDATE(1,DATE)
C CONSTANT NEEDED LATER  D1=2./9.=0.222222
      D1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=0.196854
      CC2=0.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER OF CASES TO RUN
      NC=1000
C NUMBER OF NON-SIGNIFICANT CASES - ROW,COL & TABLE
      NSIG(1)=0
      NSIG(2)=0
      NSIG(3)=0
C PAGE & LINE CNTS
      LINE=100
      IPAGE=1
      IY=9
C LOOP OVER ALL CASES
      DO 306 K=1,NC
C GENERATE CELL SIZE  (.GT. 10)
 300  IX=IY
      CALL RANDU(IX,IY,PP)
      NN=PP*100.
      IF(NN.LT.10) GO TO 300
C CALC CELL SIZES
      NSUM=0
```

```
      DO 301 I=1,3
  302 IX=IY
      CALL RANDU(IX,IY,A)
      NN=A*100.
      IF(NN.LT.10) GO TO 302
      N(I)=NN
      NSUM=NSUM+NN
  301 CONTINUE
      N4=N2*N3/N1
C IF N4 IS LESS THAN 6 DROP CASE
      IF(N4.LT.6) GO TO 300
      NSUM=NSUM+N4
      SUMN=NSUM
      DO 309 I=1,4
      AN=N(I)
      P(I)=AN/SUMN
  309 CONTINUE
C GENERATE MEMBERS OF EACH CELL & CELL SUM
  310 TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
      SUMS(I)=0.0
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
      IF(I.LE.2) A=A+5
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
  304 CONTINUE
      TSUMS=TSUMS+SUMS(I)
  303 CONTINUE
C CALC ROW,COL,& TABLE SUMS
C   CALC ROW,COL,& TABLE SUMS SO ROW FACTOR SIGNIF
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,$ TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
      FC=SSC/SERRM
      FRC=SSRC/SERRM
C IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 310
      IF(FC.LT.0.0) GO TO 310
      IF(FRC.LT.0.0) GO TO 310
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      D2=D1/DF
```

```
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIGNIFICANT CASES - ROW,COL & TABLE
      IF(PF(I).LE.0.95) NSIG(I)=NSIG(I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(6,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(6,201)P,N,F,PF
      LINE=LINE+1
  306 CONTINUE
C CALC SIGNIFICANT CASES,PERCENTAGES, & PRINT SUMMARY
      ANC=NC
      DO 308 L=1,6,2
      M=L/2+1
      LSIG=NSIG(M)
      MSIG=NC-LSIG
      PRCT(L)=FLOAT(MSIG)*100/ANC
      PRCT(L+1)=FLOAT(LSIG)*100/ANC
  308 CONTINUE
      WRITE(6,202)DATE,PRCT,NC
      STOP
      END
```

40

```
C ON120 ORAVECZ-NEWMAN HYPOTHESIS - NONSIGNIF DISPROP CASE
C MODIFIED FROM ON100  03-09-76  WD WHEATON
C DEFINITION - 2X2 CONTINGENCY TABLE
C  CELL(1,1) EQU CELL(1)
C  CELL(1,2) EQU CELL(2)
C  CELL(2,1) EQU CELL(3)
C  CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),FSIG(2,6)
C DEFN - MATRIX NSIG(2,3) CONTAINS COUNT OF NON-SIG F'S
C   COL 1 CHI SQ'D SIG
C   COL 2 CHI SQ'D NON-SIG
C   ROW 1  ROW OF TABLE
C   ROW 2  COL OF TABLE
C   ROW 3  TABLE
      INTEGER*4 N(4),SIG,NSIG(2,3),SCHI,DATE(2)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
      DATA NSIG/6*0/,LTRN,LTRS/'N','S'/
  200 FORMAT('1',35X,'ON120  ORAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES
     1  ',4A2,T123,'PAGE',I3//
     2' ' CELL SIZE RATIOS  CELL POPULATION SIZE
     3  F VALUES  PROBABILITIES'/
     4' ----------  ----------
     5 ----------  ----------'/
     6'  (1,1)  (1,2)  (2,1)  (2,2)  (1,1)  (1,2)  (2,1)  (2,2)
     7   ROW  COL  TABLE  ROW  COL  TABLE'/
     8' ------  -------  -------  -------  -----  -----  -----  -----
     9 ---------  ---------  ---------  -------  -------  -------')
  201 FORMAT(F8.5,3F9.5,1X,4I7,1X,3F12.4,1X,3F9.5,1X,A1)
  202 FORMAT('1ON120  ORAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES  ',2A4
     1///' CHI SQ''D NON-SIG  F TEST RESULTS (PERCENTS)'//
     210X,'NON-SIG  SIG  '/
     310X,'-------  -------'/
     4' ROW  ',2F11.2//
     5' COL  ',2F11.2//
     6' TABLE',2F11.2//
     7///' CHI SQ''D SIG  F TEST RESULTS (PERCENTS)'//
     810X,'NON-SIG  SIG  '/
     910X,'-------  -------'/
     A' ROW  ',2F11.2//
     B' COL  ',2F11.2//
     C' TABLE',2F11.2//
     D///' CASES CONSIDERED FOR EACH STATE CHI SQ''D =',I4)
C FIX TOTAL NUMBER OF CASES
      NC=1000
C GET CURRENT DATE
      CALL TDATE(1,DATE)
C CONSTANT NEEDED LATER  D1=2./9.=0.222222
      D1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=0.196854
      CC2=0.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER SIG & NON-SIG CASES CHI SQ'D
      NSCHI=0
      SCHI=0
C PAGE & LINE CNTS
```

Rom factor not necessarily significant

41

```
      LINE=100
      IPAGE=1
      IY=9
C GENERATE CELL SIZES (BETWEEN 10 AND 1000)
  306 NSUM=0
      DO 300 I=1,3
  301 IX=IY
      CALL RANDU(IX,IY,A)
      IF(A.LT.0.10) GO TO 301
      N(I)=A*100.
      NSUM=NSUM+N(I)
  300 CONTINUE
C GENERATE N4 SO DISPROPORTIONAL
  302 IX=IY
      CALL RANDU(IX,IY,A)
      NA=A*10.+1.
      N4=N2*N3/N1+NA
      NSUM=NSUM+N4
      SUMN=NSUM
C CALC CHI SQUARED
      CHI2A=N1*N4-N2*N3
      CHI2=SUMN*CHI2A*CHI2A/((N1+N2)*(N3+N4)*(N1+N3)*(N2+N4))
C TEST SIGNIFICANCE OF CHI SQ'D
      IF(CHI2.GT.1.320) GO TO 306
C CHI SQ NON-SIG TEST FOR NC CASES
      IF(NSCHI.EQ.NC) GO TO 312
C NOT ENOUGH  ADD THIS ONE
      LS=2
      NSCHI=NSCHI+1
      SIG=LTRN
C GENERATE MEMBERS OF EACH CELL & CELL SUM
  313 TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
      SUMS(I)=0.0
      P(I)=FLOAT(N(I))/SUMN
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
  304 CONTINUE
      TSUMS=TSUMS+SUMS(I)
  303 CONTINUE
C CALC F STATISTICS
C CALC ROW,COL,& TABLE SUMS
C   CALC ROW,COL,& TABLE SUMS SO ROW FACTOR SIGNIF
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
```

42

```
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
      FC=SSC/SERRM
      FRC=SSRC/SERRM
  500 CONTINUE
C IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 316
      IF(FC.LT.0.0) GO TO 316
      IF(FRC.LT.0.0) GO TO 316
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      D2=D1/DF
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIG CASES OF F
      IF(PF(I).LE.0.95) NSIG(LS,I)=NSIG(LS,I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(3,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(3,201)P,N,F,PF,SIG
      LINE=LINE+1
C DO WE HAVE ENOUGH CASES OF SIG AND NON-SIG CHI YET
X 312 IF(NSCHI.LT.NC) GO TO 306
C CALC & PRINT SUMMARY
C MATRIX FSIG(2,6) CONTAINS THE PERCENTS AS FOLLOWS
C    COL 1  PERCENT NON-SIG F'S
C    COL 2  PERCENT SIG F'S
C    ROW 1  ROW F NON-SIG CHI SQ'D
C    ROW 2  COL F NON-SIG CHI SQ'D
C    ROW 3 TABL F NON-SIG CHI SQ'D
C    ROW 4  ROW SIG CHI SQ'D
C    ROW 5  COL SIG CHI SQ'D
C    ROW 6  TAB SIG CHI SQ'D
  315 CN=NC
      K=2
      DO 314 I=1,6
      L=I-(I-1)/3*3
      IF(I.EQ.4) K=1
      TMP=FLOAT(NSIG(K,L))
      FSIG(1,I)=TMP/CN*100.
      FSIG(2,I)=(CN-TMP)/CN*100.
  314 CONTINUE
      WRITE(3,202)DATE,FSIG,NC
      STOP
C F(I) IS NEG DROP CASE
  316 IF(SIG.EQ.LTRS) SCHI=SCHI-1
      IF(SIG.EQ.LTRN) NSCHI=NSCHI-1
      GO TO 306
      END
```

```
& LEVEL  21                    MAIN                    DATE = 76175          10/34/32
C ON120 ORAVECZ-NEWMAN HYPOTHESIS - NONSIGNIF DISPROP CASE
C MODIFIED FROM ON100  03-09-76  WD WHEATON   ROW FACTOR SIGNIFICANT
C DEFINITION - 2X2 CONTINGENCY TABLE
C  CELL(1,1) EQU CELL(1)
C  CELL(1,2) EQU CELL(2)
C  CELL(2,1) EQU CELL(3)
C  CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),FSIG(2,6)
C DEFN - MATRIX NSIG(2,3) CONTAINS COUNT OF NON-SIG F'S
C   COL 1 CHI SQ'D SIG
C   COL 2 CHI SQ'D NON-SIG
C   ROW 1  ROW OF TABLE
C   ROW 2  COL OF TABLE
C   ROW 3  TABLE
      INTEGER*4 N(4),SIG,NSIG(2,3),SCHI,DATE(2)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
      DATA NSIG/6*0/,LTRN,LTRS/'N','S'/
  200 FORMAT('1',35X,'ON120  ORAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES
     1  ',4A2,T123,'PAGE',I3//
     2'           CELL SIZE RATIOS                    CELL POPULATION SIZE
     3                F VALUES                          PROBABILITIES'/
     4' ----------------------------------   ----------------------------
     5  ----------------------------------   ----------------------------'/
     6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
     7    ROW          COL          TABLE      ROW       COL       TABLE'/
     8' -------  -------  -------  -------   -----  -----  -----  -----
     9  ----------  ----------  ----------   -------  -------  -------')
  201 FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5,1X,A1)
  202 FORMAT('1ON120  ORAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES  ',2A4
     1///' CHI SQ''D NON-SIG  F TEST RESULTS (PERCENTS)'//
     210X,'NON-SIG      SIG  '/
     310X,'-------    -------'/
     4' ROW  ',2F11.2//
     5' COL  ',2F11.2//
     6' TABLE',2F11.2//
     7///' CHI SQ''D SIG  F TEST RESULTS (PERCENTS)'//
     810X,'NON-SIG      SIG  '/
     910X,'-------    -------'/
     A' ROW  ',2F11.2//
     B' COL  ',2F11.2//
     C' TABLE',2F11.2//
     D///' CASES CONSIDERED FOR EACH STATE CHI SQ''D =',I4)
C FIX TOTAL NUMBER OF CASES
      NC=1000
C GET CURRENT DATE
      CALL TDATE(1,DATE)
C CONSTANT NEEDED LATER  C1=2./9.=0.222222
      C1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=0.196854
      CC2=0.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER SIG & NON-SIG CASES CHI SQ'D
      NSCHI=0
      SCHI=0
C PAGE & LINE CNTS
```

```
      LINE=100
      IPAGE=1
      IY=9
C GENERATE CELL SIZES (BETWEEN 10 AND 1000)
 306  NSUM=0
      DO 300 I=1,3
 301  IX=IY
      CALL RANDU(IX,IY,A)
      IF(A.LT.0.10) GO TO 301
      N(I)=A*100.
      NSUM=NSUM+N(I)
 300  CONTINUE
C GENERATE N4 SO DISPROPORTIONAL
X302  IX=IY
     X CALL RANDU(IX,IY,A)
     X NA=A*10.+1.
     X N4=N2*N3/N1+NA
     X NSUM=NSUM+N4
      SUMN=NSUM
C CALC CHI SQUARED
      CHI2A=N1*N4-N2*N3
      CHI2=SUMN*CHI2A*CHI2A/((N1+N2)*(N3+N4)*(N1+N3)*(N2+N4))
C TEST SIGNIFICANCE OF CHI SQ'D
      IF(CHI2.GT.1.320) GO TO 306
C CHI SQ NON-SIG TEST FOR NC CASES
      IF(NSCHI.EQ.NC) GO TO 312
C NOT ENOUGH  ADD THIS ONE
      LS=2
      NSCHI=NSCHI+1
      SIG=LTRN
C GENERATE MEMBERS OF EACH CELL & CELL SUM
 313  TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
       SUMS(I)=0.0
      P(I)=FLOAT(N(I))/SUMN
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
C FOR CELLS 1 AND 2 ADD FIVE TO EACH DATA POINT
      IF(I.LE.2) A=A+5
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
 304  CONTINUE
      TSUMS=TSUMS+SUMS(I)
 303  CONTINUE
C CALC F STATISTICS
C CALC ROW,COL,& TABLE SUMS
C    CALC ROW,COL,& TABLE SUMS SO ROW FACTOR SIGNIF
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
```

```
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
      FC=SSC/SERRM
      FRC=SSRC/SERRM
  500 CONTINUE
C IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 316
      IF(FC.LT.0.0) GO TO 316
      IF(FRC.LT.0.0) GO TO 316
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      D2=D1/DF
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIG CASES OF F
      IF(PF(I).LE.0.95) NSIG(LS,I)=NSIG(LS,I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(3,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(3,201)P,N,F,PF,SIG
      LINE=LINE+1
C DO WE HAVE ENOUGH CASES OF SIG AND NON-SIG CHI YET
  312 IF(NSCHI.LT.NC) GO TO 306
C CALC & PRINT SUMMARY
C MATRIX FSIG(2,6) CONTAINS THE PERCENTS AS FOLLOWS
C    COL 1  PERCENT NON-SIG F'S
C    COL 2  PERCENT SIG F'S
C    ROW 1  ROW F NON-SIG CHI SQ'D
C    ROW 2  COL F NON-SIG CHI SQ'D
C    ROW 3 TABL F NON-SIG CHI SQ'D
C    ROW 4  ROW SIG CHI SQ'D
C    ROW 5  COL SIG CHI SQ'D
C    ROW 6  TAB SIG CHI SQ'D
  315 CN=NC
      K=2
      DO 314 I=1,6
      L=I-(I-1)/3*3
      IF(I.EQ.4) K=1
      TMP=FLOAT(NSIG(K,L))
      FSIG(1,I)=TMP/CN*100.
      FSIG(2,I)=(CN-TMP)/CN*100.
  314 CONTINUE
      WRITE(3,202)DATE,FSIG,NC
      STOP
C F(I) IS NEG DROP CASE
  316 IF(SIG.EQ.LTRS) SCHI=SCHI-1
      IF(SIG.EQ.LTRN) NSCHI=NSCHI-1
      GO TO 306
      END
```

46

```
C ON120 ORAVECZ-NEWMAN HYPOTHESIS - SIGNIF DISPROPORTIONAL CASE
C MODIFIED FROM ON100  03-09-76  WD WHEATON   Row FACTOR NOT
C DEFINITION - 2X2 CONTINGENCY TABLE          NECCESSARILY SIGNIFICANT
C  CELL(1,1) EQU CELL(1)
C  CELL(1,2) EQU CELL(2)
C  CELL(2,1) EQU CELL(3)
C  CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),FSIG(2,6)
C DEFN - MATRIX NSIG(2,3) CONTAINS COUNT OF NON-SIG F'S
C   COL 1 CHI SQ'D SIG
C   COL 2 CHI SQ'D NON-SIG
C   ROW 1  ROW OF TABLE
C   ROW 2  COL OF TABLE
C   ROW 3  TABLE
      INTEGER*4 N(4),SIG,NSIG(2,3),SCHI,DATE(2)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
      DATA NSIG/6*0/,LTRN,LTRS/'N','S'/
  200 FORMAT('1',35X,'ON120  ORAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES
     1  ',4A2,T123,'PAGE',I3//
     2'          CELL SIZE RATIOS                    CELL POPULATION SIZE
     3                F VALUES                          PROBABILITIES'/
     4' ------------------------------------   ------------------------
     5 ------------------------------------   ------------------------'/
     6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
     7     ROW          COL          TABLE       ROW       COL       TABLE'/
     8' -------  -------  -------  -------   -----  -----  -----  -----
     9 ----------  ----------  ----------  -------  -------  -------')
  201 FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5,1X,A1)
  202 FORMAT('1ON120  ORAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES  ',2A4
     1///' CHI SQ''D NON-SIG  F TEST RESULTS (PERCENTS)'//
     210X,'NON-SIG      SIG  '/
     310X,'-------    -------'/
     4' ROW  ',2F11.2//
     5' COL  ',2F11.2//
     6' TABLE',2F11.2//
     7///' CHI SQ''D SIG  F TEST RESULTS (PERCENTS)'//
     810X,'NON-SIG      SIG  '/
     910X,'-------    -------'/
     A' ROW  ',2F11.2//
     8' COL  ',2F11.2//
     C' TABLE',2F11.2//
     D///' CASES CONSIDERED FOR EACH STATE CHI SQ''D =',I4)
C FIX TOTAL NUMBER OF CASES
      NC=1000
C GET CURRENT DATE
      CALL TDATE(1,DATE)
C CONSTANT NEEDED LATER  D1=2./9.=0.222222
      D1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=0.196854
      CC2=0.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER SIG & NON-SIG CASES CHI SQ'D
      NSCHI=0
      SCHI=0
C PAGE & LINE CNTS
```

```
      LINE=100
      IPAGE=1
      IY=9
C GENERATE CELL SIZES (BETWEEN 10 AND 1000)
  306 NSUM=0
    X DO 300 I=1,4
  301 IX=IY
      CALL RANDU(IX,IY,A)
      IF(A.LT.0.10) GO TO 301
      N(I)=A*1000.
      NSUM=NSUM+N(I)
  300 CONTINUE
      SUMN=NSUM
C CALC CHI SQUARED
      CHI2A=N1*N4-N2*N3
      CHI2=SUMN*CHI2A*CHI2A/((N1+N2)*(N3+N4)*(N1+N3)*(N2+N4))
C TEST SIGNIFICANCE OF CHI SQ'D
    X IF(CHI2.LE.3.841) GO TO 306
XC CHI SQ'D SIG  TEST FOR NC CASES
    X IF(SCHI.EQ.NC) GO TO 312
XC NOT ENOUGH  ADD THIS ONE
    X LS=1
    X SCHI=SCHI+1
    X SIG=LTRS
C GENERATE MEMBERS OF EACH CELL & CELL SUM
  313 TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
       SUMS(I)=0.0
      P(I)=FLOAT(N(I))/SUMN
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
  304 CONTINUE
      TSUMS=TSUMS+SUMS(I)
  303 CONTINUE
C CALC F STATISTICS
C CALC ROW,COL,& TABLE SUMS
C   CALC ROW,COL,& TABLE SUMS SO ROW FACTOR SIGNIF
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC
C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
      FR=SSR/SERRM
```

48

```
      FC=SSC/SERRM
      FRC=SSRC/SERRM
  500 CONTINUE
C IF F(I) IS NEGATIVE DROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 316
      IF(FC.LT.0.0) GO TO 316
      IF(FRC.LT.0.0) GO TO 316
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      D2=D1/DF
      DO 305 I=1,3
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIG CASES OF F
      IF(PF(I).LE.0.95) NSIG(LS,I)=NSIG(LS,I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(3,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(3,201)P,N,F,PF,SIG
      LINE=LINE+1
C DO WE HAVE ENOUGH CASES OF SIG AND NON-SIG CHI YET
X312  IF(SCHI.LT.NC) GO TO 306
C CALC & PRINT SUMMARY
C MATRIX FSIG(2,6) CONTAINS THE PERCENTS AS FOLLOWS
C    COL 1  PERCENT NON-SIG F'S
C    COL 2  PERCENT SIG F'S
C    ROW 1  ROW F NON-SIG CHI SQ'D
C    ROW 2  COL F NON-SIG CHI SQ'D
C    ROW 3 TABL F NON-SIG CHI SQ'D
C    ROW 4  ROW SIG CHI SQ'D
C    ROW 5  COL SIG CHI SQ'D
C    ROW 6  TAB SIG CHI SQ'D
  315 CN=NC
      K=2
      DO 314 I=1,6
      L=I-(I-1)/3*3
      IF(I.EQ.4) K=1
      TMP=FLOAT(NSIG(K,L))
      FSIG(1,I)=TMP/CN*100.
      FSIG(2,I)=(CN-TMP)/CN*100.
  314 CONTINUE
      WRITE(3,202)DATE,FSIG,NC
      STOP
C F(I) IS NEG DROP CASE
  316 IF(SIG.EQ.LTRS) SCHI=SCHI-1
      IF(SIG.EQ.LTRN) NSCHI=NSCHI-1
      GO TO 306
      END
```

```
LEVEL  21                    MAIN                    DATE = 77356        11/40/57

C CN120 CRAVECZ-NEWMAN HYPOTHESIS - SIGNIF DISPROPORTIONAL CASE
C MODIFIED FROM CN100  03-09-76  WD WHEATON    ROW FACTOR SIGNIFICANT
C DEFINITION - 2X2 CONTINGENCY TABLE
C  CELL(1,1) EQU CELL(1)
C  CELL(1,2) EQU CELL(2)
C  CELL(2,1) EQU CELL(3)
C  CELL(2,2) EQU CELL(4)
      REAL*4 P(4),SUM(4),SUMS(4),F(3),PF(3),FSIG(2,6)
C DEFN - MATRIX NSIG(2,3) CONTAINS COUNT OF NON-SIG F'S
C   COL 1 CHI SQ'D SIG
C   COL 2 CHI SQ'D NON-SIG
C   ROW 1  ROW OF TABLE
C   ROW 2  COL OF TABLE
C   ROW 3  TABLE
      INTEGER*4 N(4),SIG,NSIG(2,3),SCHI,DATE(2)
      EQUIVALENCE (SUM(1),SUM1),(SUM(2),SUM2),(SUM(3),SUM3),
     1(SUM(4),SUM4),(N(1),N1),(N(2),N2),(N(3),N3),(N(4),N4),
     2(F(1),FR),(F(2),FC),(F(3),FRC)
      DATA NSIG/6*0/,LTRA,LTRS/'N','S'/
  200 FORMAT('1',35X,'CN120  CRAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES
     1  ',4A2,T123,'PAGE',I3//
     2'           CELL SIZE RATIOS                  CELL POPULATION SIZE
     3                F VALUES                       PROBABILITIES'/
     4' ------------------------------------   ------------------------
     5  ------------------------------------   ------------------------'/
     6'  (1,1)    (1,2)    (2,1)    (2,2)    (1,1)  (1,2)  (2,1)  (2,2)
     7     ROW         COL         TABLE       ROW       COL       TABLE'/
     8' -------  -------  -------  -------  -------  -------  -------  -------
     9  ----------  ----------  ----------  -------  -------  -------')
  201 FORMAT(F8.5,3F9.5,1X,4I7,1X,3E12.4,1X,3F9.5,1X,A1)
  202 FORMAT('1CN120  CRAVECZ & NEWMAN DISPROPORTIONAL CELL SIZES  ',2A4
     1///' CHI SQ''D NON-SIG  F TEST RESULTS (PERCENTS)'//
     210X,'NON-SIG      SIG  '/
     310X,'-------    -------'/
     4' ROW  ',2F11.2//
     5' COL  ',2F11.2//
     6' TABLE',2F11.2//
     7///' CHI SQ''D SIG  F TEST RESULTS (PERCENTS)'//
     810X,'NON-SIG      SIG  '/
     910X,'-------    -------'/
     A' ROW  ',2F11.2//
     B' COL  ',2F11.2//
     C' TABLE',2F11.2//
     D///' CASES CONSIDERED FOR EACH STATE CHI SQ''D =',I4)
C FIX TOTAL NUMBER OF CASES
      NC=2
C FIX NUMBER OF CASES IN DEBUG
      ICNT=0
C GET CURRENT DATE
      CALL DATE(1,DATE)
C CONSTANT NEEDED LATER  C1=2./9.=0.222222
      C1=0.222222
C PROBABILITY EXPANSION CONSTANTS
      CC1=1.196854
      CC2=1.115194
      CC3=0.000344
      CC4=0.019527
C NUMBER SIG & NON-SIG CASES CHI SQ'D
      NSCHI=0
```

```
      LINE=100
      IPAGE=1
      IY=5
C GENERATE CELL SIZES (BETWEEN 10 AND 1000)
  306 NSUM=0
      DO 300 I=1,4
  301 IX=IY
      CALL RANDU(IX,IY,A)
      IF(A.LT..10) GO TO 301
      N(I)=A*1000.
      NSUM=NSUM+N(I)
  300 CONTINUE
      SUMN=NSUM
C CALC CHI SQUARED
      CHI2A=N1*N4-N2*N3
      CHI2=SUMN*CHI2A*CHI2A/((N1+N2)*(N3+N4)*(N1+N3)*(N2+N4))
C TEST SIGNIFICANCE OF CHI SQ'D
      IF(CHI2.LE.3.841) GO TO 306
C CHI SQ'D SIG  TEST FOR NC CASES
      IF(SCHI.EQ.NC) GO TO 312
C NOT ENOUGH  ADD THIS ONE
      LS=1
      SCHI=SCHI+1
      SIG=LTRS
C GENERATE MEMBERS OF EACH CELL & CELL SUM
  313 TSUMS=0.0
      DO 303 I=1,4
      SUM(I)=0.0
      SUMS(I)=0.0
      P(I)=FLOAT(N(I))/SUMN
      L=N(I)
      DO 304 J=1,L
      IX=IY
      CALL RANDU(IX,IY,A)
C FOR CELLS 1 AND 2 ADD FIVE TO EACH DATA POINT
      IF(I.LE.2) A=A+5
      SUM(I)=SUM(I)+A
      SUMS(I)=SUMS(I)+A*A
  304 CONTINUE
      TSUMS=TSUMS+SUMS(I)
  303 CONTINUE
C CALC F STATISTICS
C CALC ROW,COL,& TABLE SUMS
C   CALC ROW,COL,& TABLE SUMS SC ROW FACTOR SIGNIF
      R1S=SUM1+SUM2
      R2S=SUM3+SUM4
      C1S=SUM1+SUM3
      C2S=SUM2+SUM4
      TABSUM=R1S+R2S
C CALC ROW,COL,& TABLE SUMS SQUARED
      TSS=TABSUM*TABSUM/FLOAT(NSUM)
      SSR=R1S*R1S/(N1+N2)+R2S*R2S/(N3+N4)-TSS
      SSC=C1S*C1S/(N1+N3)+C2S*C2S/(N2+N4)-TSS
      SST=TSUMS-TSS
      SSCELL=SUM1*SUM1/N1+SUM2*SUM2/N2+SUM3*SUM3/N3+SUM4*SUM4/N4-TSS
      SSRC=SSCELL-SSR-SSC
      SSERR=SST-SSR-SSC-SSRC

LEVEL  21                    MAIN                DATE = 77056          11/40/57

C CALC F VALUES FOR ROWS,COLS,& TABLE
      DF=NSUM-4
      SERRM=SSERR/DF
```

```
      FC=SSC/SERRM
      FRC=SSRC/SERRM
  304 CONTINUE
C IF F(I) IS NEGATIVE CROP CASE FROM CONSIDERATION
      IF(FR.LT.0.0) GO TO 316
      IF(FC.LT.0.0) GO TO 316
      IF(FRC.LT.0.0) GO TO 316
C CALC PROBABILITIES FOR ROWS,COLS & TABLE
      C2=C1/DF
      DO 305 I=1,2
      X1=SQRT(F(I))*(1.-0.25/DF)/SQRT(1.+0.5*F(I)/DF)
      X2=1.0+X1*(CC1+X1*(CC2+X1*(CC3+X1*CC4)))
      X3=X2*X2*X2*X2
      PF(I)=1.-1./X3
C CNT NON-SIG CASES OF F
      IF(PF(I).LE.0.95) NSIG(LS,I)=NSIG(LS,I)+1
  305 CONTINUE
C PRINT CASE PARAMETERS
      IF(LINE.LE.56) GO TO 307
      WRITE(3,200)DATE,IPAGE
      LINE=6
      IPAGE=IPAGE+1
  307 WRITE(3,201)P,N,F,PF,SIG
      LINE=LINE+1
C DO WE HAVE ENOUGH CASES OF SIG AND NON-SIG CHI YET
  312 IF(SCHI.LT.NC) GO TO 306
C CALC & PRINT SUMMARY
C MATRIX FSIG(2,6) CONTAINS THE PERCENTS AS FOLLOWS
C   COL 1  PERCENT NON-SIG F'S
C   COL 2  PERCENT SIG F'S
C   ROW 1  ROW F NON-SIG CHI SQ'D
C   ROW 2  COL F NON-SIG CHI SQ'D
C   ROW 3  TABL F NON-SIG CHI SQ'D
C   ROW 4  ROW SIG CHI SQ'D
C   ROW 5  COL SIG CHI SQ'D
C   ROW 6  TAB SIG CHI SQ'D
  315 CN=NC
      K=2
      DO 314 I=1,6
      L=I-(I-1)/3*3
      IF(I.EQ.4) K=1
      TMP=FLOAT(NSIG(K,L))
      FSIG(I,I)=TMP/CN*100.
      FSIG(2,I)=(CN-TMP)/CN*100.
  314 CONTINUE
      WRITE(3,202)DATE,FSIG,NC
      STOP
C F(I) IS NEG CROP CASE
  316 IF(SIG.EQ.LTRS) SCHI=SCHI-1
      IF(SIG.EQ.LTRN) NSCHI=NSCHI-1
      GO TO 306
      OVRLG UNIT(3)
      AT 304
      DISPLAY I,A
      AT 506

LEVEL  21              MAIN              DATE = 77056          11/40/57

      DISPLAY TSUMS,R1S,R2S,C1S,C2S,TABSUM,TSS,SSR,SSC,SST,SSCELL,SSRC,
     1SSERR,DF,SERRM,FR,FC,FRC
      ICNT=ICNT+1
      IF(ICNT.GT.2) STOP
      END
```